AFRL-AFOSR-JP-TR-2017-0047

Decomposing Huge Networks into Skeleton Graphs by Reachable Relations

Kazumi Saito
University Of Shizuoka

06/07/2017
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/ IOA
Arlington, Virginia 22203
Air Force Materiel Command

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
07-06-2017

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
23 May 2016 to 22 May 2017

4. TITLE AND SUBTITLE
Decomposing Huge Networks into Skeleton Graphs by Reachable Relations

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA2386-16-1-4032

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Kazumi Saito

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University Of Shizuoka
52-1 Yada, Suruga-ku
Shizuoka, 422-8526 JP

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AOARD
UNIT 45002
APO AP 96338-5002

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR IOA

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-JP-TR-2017-0047

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The research team developed new techniques for accelerating critical link detection and distance-based centrality computation by decomposing huge networks into skeleton graphs by reachable relations. The main research results are in three new approaches, 1) efficient detection of critical links in a large network by using bottom-k sketch algorithm and by employing two new acceleration techniques: marginal-link updating and redundant-link skipping, 2) accurate and efficient detection of such critical links by a new method which consists of one existing and two new acceleration techniques: redundant-link skipping, marginal-node pruning and burn-out following, and 3) accelerating computation of distance-based closeness and betweenness centrality measures by pruning some nodes and links based on the cut links of a given spatial network.

15. SUBJECT TERMS
Graph node reachability, Strongly connected component, Quotient graph, Skeleton graph, Node pruning

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
Unclassified	Unclassified	Unclassified	SAR	12

19a. NAME OF RESPONSIBLE PERSON
KNOPP, JEREMY

19b. TELEPHONE NUMBER *(Include area code)*
315-227-7006

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18

Final Report for AOARD Grant FA2386-16-1-4032

"Decomposing Huge Networks into Skeleton Graphs by Reachable Relations"

22/05/201

Name of Principal Investigators (PI and Co-PIs): Kazumi Saito

- e-mail address : k-saito@u-shizuoka-ken.ac.jp
- Institution : School of Administration and Informatics, University of Shizuoka
- Mailing Address : 52-1 Yada, Suruga-ku, Shizuoka 422-8526 Japan
- Phone : +81-54-264-5436
- Fax : +81-54-264-5436

Period of Performance: 05/23/2016– 05/22/2017

Abstract:

We developed new techniques for accelerating critical link detection and distance-based centrality computation by decomposing huge networks into skeleton graphs by reachable relations. The main research results are on three new approaches, 1) efficient detection of critical links in a large network by using bottom-*k* sketch algorithm and further by employing two new acceleration techniques: marginal-link updating and redundant-link skipping (Saito, et al. 2016), 2) accurate and efficient detection of such critical links by proposing a new method which consists of one existing and two new acceleration techniques: redundant-link skipping, marginal-node pruning and burn-out following (Saito, et al. 2017), and 3) accelerating computation of distance-based closeness and betweenness centrality measures by pruning some nodes and links based on the cut links of a given spatial network (Ohara, et al. 2016).

First, we addressed the problem of efficiently detecting critical links in a large network. Critical links are such links that their deletion exerts substantial effects on the network performance. In our research, we defined the performance as being the average node reachability. We tackled this problem by using bottom-*k* sketch algorithm and further by employing two new acceleration techniques: marginal-link updating (MLU) and redundant-link skipping & pruning (RLS), where RLS decomposes huge networks into skeleton graphs by pruning redundant links. We tested the effectiveness of the proposed method using two real-world large networks and two synthetic large networks and showed that the new method can compute the performance degradation by link removal about an order of magnitude faster than the baseline method in which bottom-*k* sketch algorithm is applied directly. Further, we confirmed that the measures easily composed by well-known existing centralities, e.g. in/out-degree, betweenness, PageRank, authority/hub, are not able to detect critical links. Links detected by these measures do not reduce the average reachability at all, i.e. not critical at all.

Second, we also tackled the same problem of efficiently detecting critical links in a large network problem by proposing a new method which consists of one existing and two new acceleration techniques: redundant-link skipping & pruning (RLS), marginal-node pruning (MNP) and burn-out following (BOF), where RLS and MNP decompose huge networks into skeleton graphs by pruning redundant links and nodes. All of them are designed to avoid unnecessary computation and work both in combination and in isolation. We also tested the effectiveness of the proposed method using two real-world large networks and two synthetic large networks. In particular, we showed that the new method can compute the performance degradation by link removal without introducing any approximation within a comparable computation time needed by the bottom-*k* sketch which

is a summary of dataset and can efficiently process approximate queries, i.e., reachable nodes, on the original dataset, i.e., the given network.

Finally, by focusing on spatial networks embedded in the real space, we extended the conventional step-based closeness and betweenness centralities by incorporating inter-nodes link distances obtained from the positions of nodes. Then, we proposed a method for accelerating computation of these centrality measures by pruning some nodes and links based on the cut links of a given spatial network, which performs a decomposition into its skeleton graph. In our experiments using spatial networks constructed from urban streets of cities of several types, our proposed method achieved about twice the computational efficiency compared with the baseline method. Actual amount of reduction in computation time depends on network structures. We further experimentally showed by examining the highly ranked nodes that the closeness and betweenness centralities have completely different characteristics to each other.

Introduction:

Studies of the structure and functions of large networks have attracted a great deal of attention in many different fields of science and engineering (Newman 2003). Developing new methods/tools that enable us to quantify the importance of each individual node and link in a network is crucially important in pursuing fundamental network analysis. Networks mediate the spread of information, and it sometimes happens that a small initial seed cascades to affect large portions of networks (Watts 2002). Such information cascade phenomena are observed in many situations: for example, cascading failures can occur in power grids (e.g., the August 10, 1996 accident in the western US power grid), diseases can spread over networks of contacts between individuals, innovations and rumors can propagate through social networks, and large grass-roots social movements can begin in the absence of centralized control (e.g., the Arab Spring). These problems have mostly been studied from the view point of identifying influential nodes under some assumed information diffusion model. There are other studies on identifying influential links to prevent the spread of undesirable things. We study this problem from a slightly different angle in a more general setting, that is to answer “Which links are most critical in maintaining a desired network performance?”. For example, when the desired performance is to minimize contamination, the problem is reduced to detecting critical links to remove or block. If the desired performance is to maximize evacuation or minimize isolation, the problem is to detect critical links that reduce the overall performance if these links do not function. This problem is mathematically formulated as an optimization problem when a network structure is given and a performance measure is defined. In our research, we define the performance to be the average node reachability with respect to a link deletion, i.e. average number of nodes that are reachable from every single node when a particular link is deleted/blocked. The problem is to rank the links in accordance with the performance and identify the most critical link(s).

One common approach to analyze large complex networks is investigating their characteristics through a measure called centrality. Various kinds of centralities are used according to what we want to know. For example, if our goal is to know the topological characteristics of a network, degree, closeness, and betweenness centralities (Wasserman & Faust 1994) can be used. If it is to know the importance of nodes that constitute a network, HITS and PageRank (Langville & Meyer 2005) centralities are often used. Influence degree centrality (Kimura, et al. 2016) is another one to measure the importance of nodes. Among these conventional centralities, we focus on the closeness and betweenness centralities in this work because they are closely related to real world problems such as location planning of commercial or evacuation facilities in a wide area. Here, note that closeness and betweenness centralities usually approximate the distance between two distinct nodes by

the number of links traversed to get to one node from another. This approximation may not be realistic when analyzing networks such as real traffic networks, one of the real world problems. Thus, as a particular class, we focus on spatial networks embedded in the real space, like urban streets, whose nodes occupy a precise position in two or three-dimensional Euclidean space, and whose links are real physical connections (Crucitti, et al. 2006). Analyzing and characterizing the structure of such large spatial networks will play an important role for understanding and improving the usages of these networks, as well as discovering new insights for developing and planning city promotion, trip tours and so on. To facilitate such research work, we proposed techniques useful to accelerate their computations based on network pruning. This is motivated by the fact that the computation time to calculate the value of such conventional step-based centralities for every single node in a network becomes larger as the size of the network gets larger because of the necessity to traverse each link in the network multiple times.

Method / Theory:

Critical link detection problem: Let $G = (V, E)$ be a given simple directed network without self-loops, where $V = \{u, v, w, \ldots\}$ and $E = \{e, f, g, \ldots\}$ are sets of nodes and directed links, respectively. Each link e is also expressed as a pair of nodes, i.e., $e = (u, v)$. Below we denote the numbers of nodes and links by $N = |V|$ and $M = |E|$, respectively. Let $R(v; G)$ and $Q(v; G)$ be the sets of reachable nodes by forwardly and reversely following links from a node v over G, respectively, where note that $v \in R(v; G)$ and $v \in Q(v; G)$. Also, let $R_1(v; G)$ and $Q_1(v; G)$ be the sets of those nodes adjacent to v, i.e., $R_1(v; G) = \{w \in R(v; G) \mid (v, w) \in E\}$ and $Q_1(v; G) = \{u \in Q(v; G) \mid (u, v) \in E\}$, respectively. Now, let $G_e = (V, E \setminus \{e\})$ be the network obtained after removing a link $e = (v, w)$, then we can define the reachability degradation value with respect to $e \in E$ as follows:

$$F(e; G) = \Sigma_{x \in V} (|R(x; \mathrm{G})| - |R(x; G_e)|)/N.$$

In our research, we focus on the problem of accurately and efficiently calculating $F(e; G)$ for every $e \in E$. Of course, network performance measure is not unique. It varies from problem to problem, but computing $R(v; G_e)$ for every node $v \in V$ can be a fundamental task. Note that our proposed method and techniques can directly contribute to this task.

Approximation methods for critical link detection: Network performance varies with specific problem, but in general it is represented by the reachability performance, i.e., how many nodes are reachable from a node in the network on the average. This brings in computational issue because reachability must be estimated for all the nodes for a particular link removal and to find critical links this has to be repeated for all the links. The number of links is generally an order of magnitude larger than the number of nodes even for a sparse network that is encountered in actual practice. We used bottom-k sketch algorithm as a basis to count reachable nodes, which only uses k-samples to estimate the reachable nodes from a selected node. It has a sound theoretical background and been shown quite efficient and accurate for a k which is far smaller than the number of nodes in the network. Our contribution is to introduce two new acceleration techniques to further reduce the bottom-k sketch computation by clever local update and redundant computations pruning. The first technique MLU (marginal-link updating) locally updates the bottom-k sketches of some nodes when removing links incident to a node with in-degree 0 or out-degree 0 in the network. The second technique RLS (redundant-link skipping) selects each link that does not affect the performance with respect to its removal and prune some subset of such links.

In our proposed method referred to as the BKS method, the RLS technique is applied before the MLU techniques, because it is naturally conceivable that the RLS technique decreases the number of links in our network. Clearly we can individually incorporate these techniques into the baseline method. Hereafter, we refer to the baseline method the BL method, the BKS method without the MLU technique as the RLS method, and the BKS method without the RLS technique as the MLU method.

Exact methods for critical link detection: We explored an exact method to compute the reachability. Our contribution is that we 1) introduced three acceleration techniques to reduce redundant computation 2) evaluated the computational efficiency by comparing with the bottom-k sketch, and 3) evaluated the accuracy of bottom-k sketch and showed that it does not necessarily result in good accuracy. The first acceleration technique RLS (redundant-link skipping & pruning), which is also employed in the BKS method, selects each link that does not affect the performance with respect to its removal and prune some subset of such links. The second technique MNP (marginal node pruning) recursively performs pruning every node that has degree 1 such that its in- and out-degrees are 1 and 0 or 0 and 1, respectively. The third technique BOF (burn out following) reduces following the same link multiple times by first computing the reachable nodes from the node connected to a link to be removed (burning out) and then computing the reachable nodes by only following the nodes uniquely reachable from a given node.

In our proposed method referred to as the PM method, we apply the RLS, MNP and BOF techniques to the baseline method in this order, since it is naturally conceivable that the RLS and MNP techniques decrease the numbers of links and nodes in our network. Clearly, we can individually incorporate these techniques into the baseline method. Same as we did before, we refer to the proposed method without the RLS technique as the \RLS method, the method without the MNP technique as the \MNP method, and the method without the BOF technique as the \BOF method.

Distance based centrality computation: Let $G = (V, E)$ be a spatial network consisting of a single connected component without self-loops, where $V = \{u, v, w, \ldots\}$ and $E = \{e, f, g, \ldots\}$ are sets of nodes and undirected links, respectively. For each link $e = (u, v)$, we express the distance between nodes u and v by $d(u, v)$, where we can obtain these distances from the positions of nodes in the spatial network. For each pair of nodes $u, w \in V$ without the direct connection, we define the distance $d(u, w)$ as the geodesic distance over the network, as usual. Then, for each node $u \in V$, we can define the following distance based closeness centrality measure:

$$DC(u) = (\Sigma_{w \in V} d(u, w))^{-1}$$

Note that the distance based closeness centrality $DC(u)$ is a natural extension to the conventional step based closeness centrality $SC(u)$ because $DC(u)$ reduces to $SC(u)$ by setting $d(u, v) = 1$ for each link $(u, v) \in E$. Similarly, for each node $v \in V$, we can define the following distance based betweenness centrality measure:

$$DB(v) = \Sigma_{u \in V\setminus\{v\}} \Sigma_{w \in V\setminus\{u, v\}} \sigma(u, w; v)/\sigma(u, w)$$

where $\sigma(u, w)$ is the total number of the paths with the smallest distance between node u and node w in G and $\sigma(u, w; v)$ is the number of those paths between node u and node w in G that passes through node v. Again, note that the distance based betweenness centrality $DB(v)$ is a natural extension to the conventional step based betweenness centrality $SB(v)$ because $DB(v)$ also reduces to $SB(v)$ by setting $d(u, v) = 1$ for each link $(u, v) \in E$. By applying the best-first search algorithm starting from each node $u \in V$ with respect to distance $d(u, w)$, we can calculate these centrality measures, $DC(u)$ and $DB(v)$, for all the nodes in G. As mentioned earlier, when calculating closeness and betweenness centrality measures, their computation becomes harder as the network size increases. Below we propose a method of improving the computational efficiency to calculate these centrality measures.

Accelerating distance based centrality computation: We first extended the conventional step-based closeness and betweenness centralities to analyze spatial networks. Unlike these conventional centralities that adopt the number of links to be traversed to reach one node from another as the distance between them, the extended distance-based closeness and betweenness centralities take into account the inter-node link distances obtained from the positions of nodes. They are natural extensions of the conventional centralities and general enough to include their definitions as a special case. Second, we

have proposed two novel techniques to improve the computational efficiency to compute the distance-based centralities. Both are based on graph cut and recursively applied to a network in order to reduce scan size needed for computing the centralities. The TP (Top-down pruning) technique recursively decomposes a network into two disjoint sub-networks that are connected by a cut link, while the BP (Bottom-up pruning) technique recursively removes a degree-one node by eliminating a cut link adjacent to it before the TP technique is applied. Current algorithm is designed for undirected networks, but it is straightforward to extend these techniques so that they can deal with directed networks.

We empirically evaluated the computational efficiency of these three methods in comparison to the baseline method without the proposed pruning techniques, where the baseline method is also referred to as the BL method. In our proposed method referred to as the PR method, the BP technique is applied before the TP techniques, because we can easily know the degree-one nodes in our network. Although we can individually incorporate these techniques into the baseline method that do not employ our proposed pruning techniques, we only consider the proposed method without the TP technique, which is referred to as the BP method.

Experiments:

Data sets for critical link detection: Using two benchmark and two synthetic networks, we evaluated the effectiveness of the proposed methods for the problem of detecting critical links. Namely, we employed two benchmark networks obtained from SNAP (Stanford Network Analysis Project)[1] The first one is a high-energy physics citation network from the e-print arXiv[2], which covers all the citations within a dataset of 34,546 papers (nodes) with 421,578 citations (links). If a paper u cites paper v, the network contains a directed link from u to v. The second one is a sequence of snapshots of the Gnutella peer-to-peer file sharing network from August 2002[3]. There are total of 9 snapshots of Gnutella network collected in August 2002. The network consists of 36,682 nodes and 88,328 directed links, where nodes represent hosts in the Gnutella network topology and links represent connections between the Gnutella hosts. In addition, we utilized two synthetic networks (around 35,000 nodes and 350,000 links) with a DAG (Directed Acyclic Graph) property, which were generated by using the DCNN and DBA methods described in (Kimura, et al. 2016), respectively. Here, networks generated by DCNN have both the small-world and scale-free properties, while those by DBA have only the scale-free property.

Data sets for centrality computation: We used OSM (OpenStreetMap) data of eight cities in our experiments, i.e., Barcelona (Spain, Europe), Bologna (Italy, Europe), Brasilia (Brazil, South America), Cairo (Egypt, Africa), Washington D.C. (United States, North America), New Delhi (India, Asia), Richmond (United States, North America), and San Francisco (United States, North America). These are a subset of cities studied in (Crucitti, et al. 2006). We obtained the OSM data of these eight cities from Metro Extracts[4] in August, 2015. Here note that in our experiments, the area of each city is more than 100 times larger than those of the previous study (Crucitti, et al. 2006). From the OSM data of each city, we extracted all highways and all nodes appearing in them, and constructed each spatial network by mapping the ends, intersections and curve-fitting-points of streets into nodes and the streets between the nodes into links. Then, based on GRS80 (Moritz, 200), we calculated each inter-node link distance from the positions of the nodes, each of which is described by a pair of latitude and longitude.

[1] https://snap.stanford.edu/
[2] https://snap.stanford.edu/data/cit-HepPh.html
[3] https://snap.stanford.edu/data/p2p-Gnutella30.html
[4] https://mapzen.com/data/metro-extracts

Results and Discussion:

Approximation methods for critical link detection: We evaluated the efficiency of the proposed method which calculates $F(e; G)$ for each link $e \in E$. We compared the computation time of the baseline (BL), RLS, MLU, and proposed (BKS) methods by performing five trials. Here, we used the same random value assignment for each trial so that the bottom-k sketches of all the nodes are the same for any method, i.e., it is guaranteed that each method can produce the same result. Figure 1 shows the computation times of each method for five trials plotted by dots and the average values over these trials plotted by different markers as indicated in the figure, where we set $k = 2^6$ for calculation of the bottom-k sketches of all the nodes. Figure 1(a) compares the actual processing times of these methods, where our programs implemented in C were executed on a computer system equipped with two Xeon X5690 3.47GHz CPUs and a 192GB main memory with a single thread within the memory capacity. Figure 1(b) compares the reduction rates of computation times for these methods from the BL method.

From Fig. 1(a), we can see that the computation times were improved largely for the CIT and CNN networks, modestly for the P2P network, and much less modestly for the DBA network, although the computation time of the BL method for the P2P network was smaller than those for the other networks. More specifically, as expected, we consider that the RLS technique worked quite well especially for the CIT and DCN networks, due to large numbers of skippable and prunable links in these networks. On the other hand, although the MLU technique is not so remarkably effective, we consider that this technique can steadily improve the reduction rate of computation times especially for the P2P network as shown in Fig. 1(b). In short, we can conjecture that the proposed method combining both the RLS and MLU techniques is more reliable than the other three methods in terms of computation time because it produced the best performance for all of the four networks. Reduction of computation time depends on network structures, but overall we can say that use of both techniques can increase the computational efficiency by about an order of magnitude. These results demonstrate the effectiveness of the proposed method.

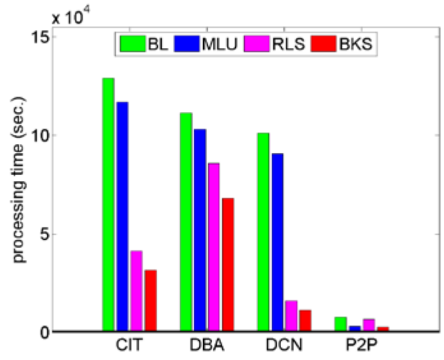

Figure 1(a): Processing times of approximation methods for critical link detection

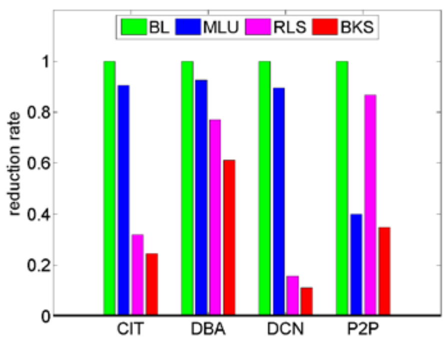

Figure 1(b): Reduction rates of approximation methods for critical link detection

Exact methods for critical link detection: We evaluated the efficiency of the proposed acceleration techniques by comparing the computation times of the \BOF, \MNP, \RLS, and the proposed (PM) methods. Figure 2 shows our experimental results which compares the actual processing times of these methods. From Fig. 2, we can clearly see that except for the DCN network, the \BOF method required much computation times compared with the other three methods. As described earlier, these experimental results can be

naturally explained from our conjecture that the \ BOF method would work well for the DCN network. We can also see that the \ MNU method exhibited the worst performance for the P2P network, while the \ RLS for the DCN network. Which technique works best depends on the network characteristics. Overall BOF which was newly introduced in this paper works the best. MNP and RLS are similar and work less. The proposed method PM combining all the three techniques BOF, MNP and RLS is most reliable and produces the best performance, but the actual reduction of computation time depends on network structure. These results demonstrate the effectiveness of the proposed method.

Figure 3 shows our experimental results by setting the parameter k of the baseline BKS method to 2^9 and the revised BKS method to 2^9 and 2^{10}, denoted as b:2^9, r:2^9, and r:2^{10}, respectively. From these results, we can see that the proposed method substantially outperforms both the baseline BKS and the revised BKS methods for the DCN network. For the other networks, it is better than the baseline BKS method for k = 2^9 and the revised BKS method for 2^{10}. Below we will see that setting k at 2^{10} is not large enough to attain a good accuracy especially for the P2P network. We can say that the proposed method is competitive to the approximation method in terms of computation efficiency and has a merit of computing the correct values for reachability degradation. The exact solutions obtained by our method can be used as the ground-truth for evaluating the approximation method. Let $E(m)$ be the set of the top-m links according to $F(e; G)$.

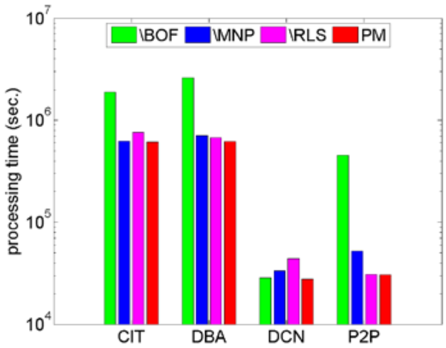

Figure 2: Processing times of exact methods for critical link detection

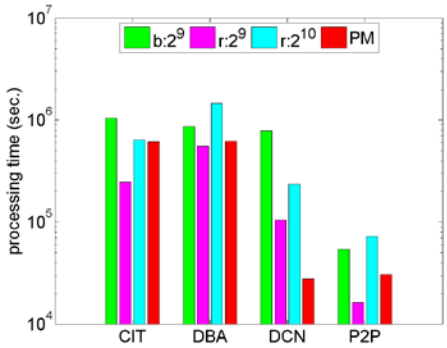

Figure 3: Processing times of approximation and exact methods for critical link detection

Figure 4 shows the average relative error of the estimated value $J(e; G)$ by BKS over $E(5)$, i.e., $\sum_{e \in E(5)} |1 - J(e; G)/F(e; G)|/5$, where we set k to one of {2^7, 2^8, 2^9, 2^{10}} for each network. From these experimental results, we observe that quite accurate estimation results were obtained for the CIT network, and the relative errors decreased monotonically by using a larger k. If we request the relative error to be less than 0.01, we need the parameter settings greater than $k = 2^8$. For other networks we observe that the results of the DBA and DCN networks are somewhat accurate around 0.1 when $k = 2^{10}$, but the results of the P2P network were quite inaccurate. We need much larger k and the computation time for BKS will overly exceed that of the present method.

We discuss below why the BKS method worked very poorly for the P2P network. As a typical situation for a given removed link $e = (u, v) \in E$, assume that $R(w; G_e) \cap R(v; G_e) = \emptyset$ for any $w \in Q(u; G_e)$, then, we obtain $Q(u; G) = Q(u; G_e)$, $R(v; G) = R(v; G_e)$, and the reachability degradation value $F(e; G) = |Q(u; G)| \times |R(v; G)|/|V|$. However, when $|R(u; G)| \approx |V|$, the BKS method may widely underestimate $J(e; G)$ if $R(v; G)$ contains a very small random value assigned by bottom-k sketch. This situation is likely to occur when $|R(u; G)| \approx |V|$ and $|R(v; G)|$ is quite small. Figure 5 shows distributions of reachability size $|R(v; G)|$ for its rank. From this figure, we can clearly see that there exist two groups of nodes in the P2P

network, those reachable to almost all of the other nodes, just like $|R(u; G)| \approx |V|$, and those reachable to almost only themselves. These results clearly support our above explanation.

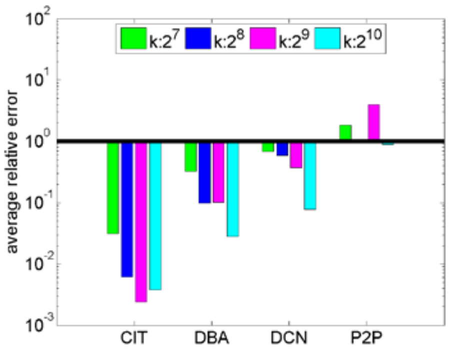

Figure 4: Relative errors of approximation method for critical link detection

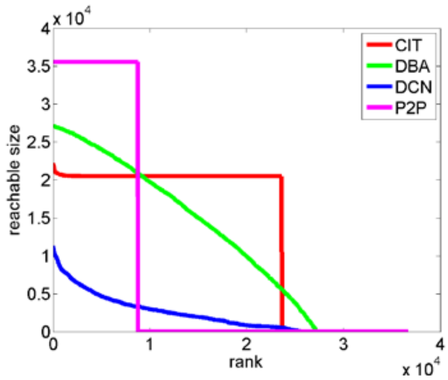

Figure 5: Reachability distributions of networks for critical link detection

Accelerating distance based centrality computation: We evaluated the efficiency of the proposed method which simultaneously calculates $DC(v)$ and $DB(v)$ for each node $v \in V$, by comparing the computation time of the baseline (BL), only bottom-up pruning (BP), and the proposed (PR) methods. We implemented the BL method based on Brandes's algorithm (Brandes, 2001) known as the standard and efficient technique for computing the betweenness centrality of each node in a network. Figure 6 shows the computation time of each method, where the abbreviations shown in the horizontal axis are Ba (Barcelona), Bo (Bologna), Br (Brasilia), Ca (Cairo), Wa (Washington D.C.), Ne (New Delhi), Ri (Richmond), Sa (San Francisco). Figure 6(a) compares the actual processing time of these methods. Figure 6(b) compares the reduction rates of computation time for these methods from the BL method. From Figs. 6(a) and 5(b), we can see that for all the networks, the BP method steadily improves the computational efficiency of the BL method, and the PR method slightly improve that of the BP method. These results demonstrate the effectiveness of the proposed techniques. More specifically, as expected, the processing time of the BL method is almost proportional to the size of network. In contrast, from Fig. 5(b), we can see that the reduction rates of the BP and PR methods depend on the network, i.e., around from 0.4 to 0.8 for the BP method and around from 0.3 to 0.7 for the PR method. These results indicate that the effects of our pruning techniques depend on the networks. On the other hand, we note that the improvement rates of the PR method over the BP method are modest, i.e., the reduction rates by the TP technique are not so remarkably effective. This must be partly because the TP technique requires additional computation costs for detecting cut links. Overall, we can conjecture that the proposed method combining both the BP and TP techniques is more reliable than the other two methods in terms of computation time because it produced the best performance for all of the eight networks. In short, reduction of computation time depends on network structures, but overall we can say that use of both techniques can increase the computational efficiency by nearly twice of the BL method.

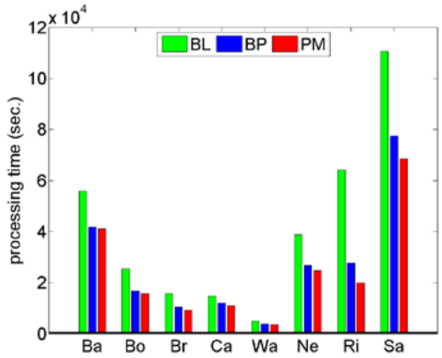

Figure 6(a): Processing time comparison for distance based centrality computation

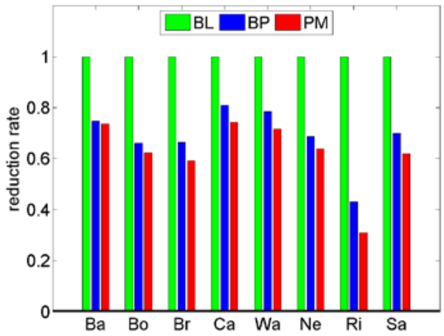

Figure 6(b): Reduction rate comparison for distance based centrality computation

References:

1. Brandes, U.: A faster algorithm for betweenness centrality. Journal of Mathematical Sociology 25, 163–177 (2001)
2. Crucitti, P., Latora, V., Porta, S.: Centrality measures in spatial networks of urban streets. Physical Review E 73(3), 036125 (2006)
3. Freeman, L.: Centrality in social networks: Conceptual clarification. Social Networks 1, 215–239 (1979)
4. Kimura, M., Saito, K., Ohara, K., Motoda, H.: Speeding-up node influence computation for huge social networks. International Journal of Data Science and Analytics 1, 1–14 (2016)
5. Langville, A.N., Meyer, C.D.: A survey of eigenvector methods for web information retrieval, SIAM review 47, pp.135-161 (2005).
6. Moritz, H.: Geodetic reference system 1980. Journal of Geodesy 74(1), 128–133 (2000)
7. Newman, M.E.J.: The structure and function of complex networks. SIAM Review 45, 167–256 (2003)
8. Ohara, K., Saito, K., Kimura, M., Motoda, H.: Accelerating computation of distance-based centrality measures for spatial networks. In: Proceedings of the 19th International Conference on Discovery Science. pp. 228–239. (2016)
9. Saito, K., Kimura, M., Ohara, K., Motoda, H.: Detecting critical links in complex network to maintain information flow/reachability. In: Proceedings of the 14th Pacific Rim International Conference on Artificial Intelligence. pp. 419–432 (2016)
10. Saito, K., Kimura, M., Ohara, K., Motoda, H.: An Accurate and Efficient method to Detect Critical Links to Maintain Information Flow in Network. In: Proceedings of the 23rd International Symposium on Methodologies for Intelligent Systems. in press, (2017)
11. Watts, D.: A simple model of global cascades on random networks. Proceedings of the National Academy of Sciences of the United States of America 99, 5766–5771 (2002)

List of Publications and Significant Collaborations that resulted from your AOARD supported project: In standard format showing authors, title, journal, issue, pages, and date, for each category list the following:

a) papers published in peer-reviewed journals,

b) papers published in peer-reviewed conference proceedings,

[1]. Kazumi Saito, Masahiro Kimura, Kouzou Ohara and Hiroshi Motoda, "Detecting Critical Links in Complex Network to Maintain Information Flow/Reachability," Proc. of the 14th Pacific Rim International Conference on Artificial Intelligence (PRICAI2016), pp.419-432, 2016.

[2]. Kazumi Saito, Kouzou Ohara, Masahiro Kimura and Hiroshi Motoda, "An Accurate and Efficient method to Detect Critical Links to Maintain Information Flow in Network," Proc. of the 23rd International Symposium on Methodologies for Intelligent Systems, (ISMIS2017), in press, 2017.

[3]. Kouzou Ohara, Kazumi Saito, Masahiro Kimura and Hiroshi Motoda, "Accelerating computation of distance-based centrality measures for spatial networks," Proc. of the 19th International Conference on Discovery Science (DS2016), pp.376-391, 2016.

c) papers published in non-peer-reviewed journals and conference proceedings,

d) conference presentations without papers,

e) manuscripts submitted but not yet published, and

f) provide a list any interactions with industry or with Air Force Research Laboratory scientists or significant collaborations that resulted from this work.

Attachments: Publications a), b) and c) listed above if possible.